

RAPPORT - TP1

IA non-supervisée

Clustering

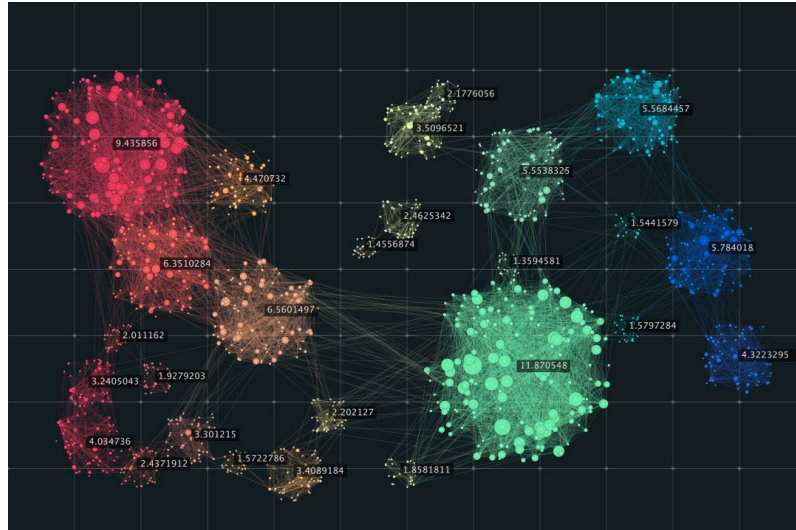


fig. 1 : Illustration Clustering

Encadrant : M. Ivan VAZQUEZ-MUNOZ
M. D. KNITTEL

10/03/2026

Question 1 : Affichage temporel des signaux

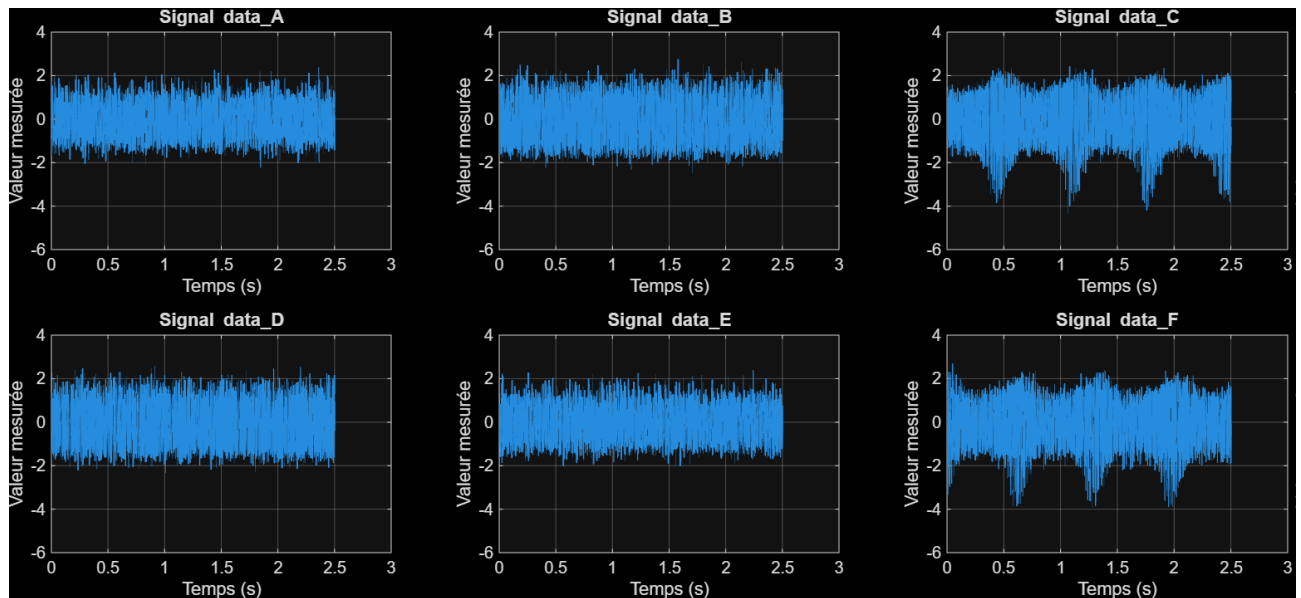


fig. 2 : Figure signaux temporels

Dans cette question, je remarque que l'affichage temporel des 6 signaux ne permet pas d'extraire beaucoup d'informations.

Seuls les signaux C et F se distinguent par des pics répétitifs, suggérant la présence d'à-coups, probablement liés à des phénomènes vibratoires ou autres au sein de la machine tournante.

Donc je vais passer en fréquentiel voire calculer des caractéristiques spécifiques tels que l'entropie de Shannon, la valeur efficace et tout ce qui suit, pour peut-être avoir plus de facilité à comprendre le comportement de ces signaux, l'objectif étant de traduire ou de retirer un maximum d'informations sur l'état des signaux de la machine.

Question 2 : Calculs des caractéristiques temporelles

Dans cette question, l'objectif est de calculer et d'analyser les neuf caractéristiques à partir des signaux data contenant les six fichiers de A à F.

Pour chacune des caractéristiques, on peut observer le comportement des six signaux comme suit :

Pour la partie code, j'ai fait un plot sur l'ensemble des signaux pour une même caractéristique donnée.

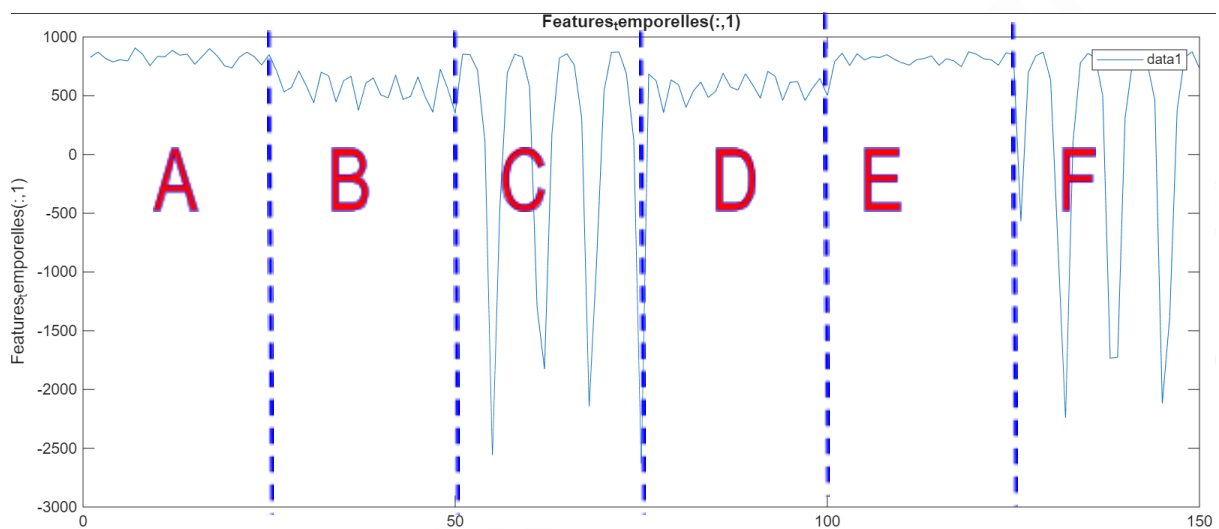


fig. 3 : Figure des caractéristiques

Pour l'entropie de Shannon, on peut distinguer respectivement les signaux A, B, C, D, E et F.

On voit qu'effectivement les signaux A et E sont très similaires, je peux donc dans un premier temps répondre à la question en disant que si le signal E est un signal normal, tout comme le signal A est normal donc sans défaut machine, je me base sur sa ressemblance au niveau de l'amplitude et sa position dans l'axe des ordonnées.

On peut aussi dire que les signaux C et F sont des signaux affectés par un défaut machine. Ainsi que les signaux B et D qui ont l'air d'être affectés par un défaut d'autre type.

Question 3 : Clustering K-means

Pour cette partie, j'ai commencé par afficher les 2 clusters (k=2) pour comprendre comment se situent les signaux A, B, C, D, E et F par rapport à leurs caractéristiques.

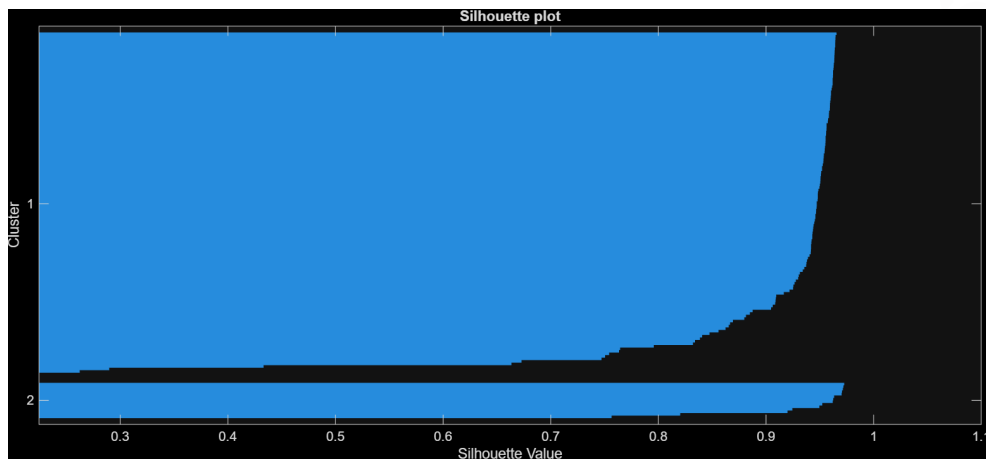


fig. 4 : Figure silhouette (k = 2=)

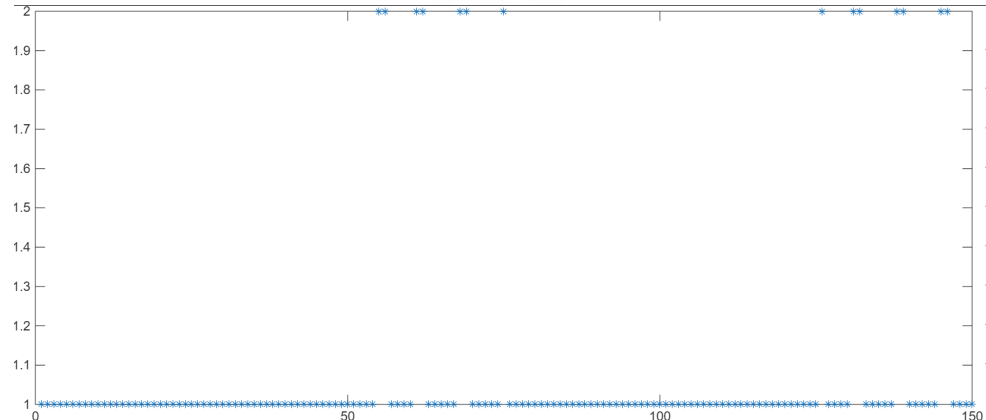


fig. 5 : Figure classification instable (k = 2)

J'ai eu des doutes sur la discontinuité au niveau du cluster 1, donc je suis passé à k = 3, pour mieux affiner mon analyse.

Avec $k = 3$, on peut encore distinguer les 6 signaux dans l'ordre respectif de gauche à droite, dont on compare encore une fois les signaux A et E qui se situent à 1, on voit que les 25 fenêtres de chacun de ces signaux sont plutôt continue.

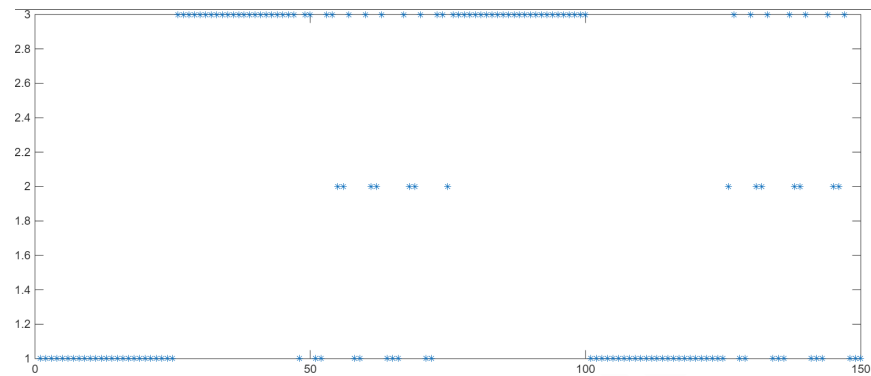


fig. 6 : Figure de la classification correcte des signaux ($k = 3$)

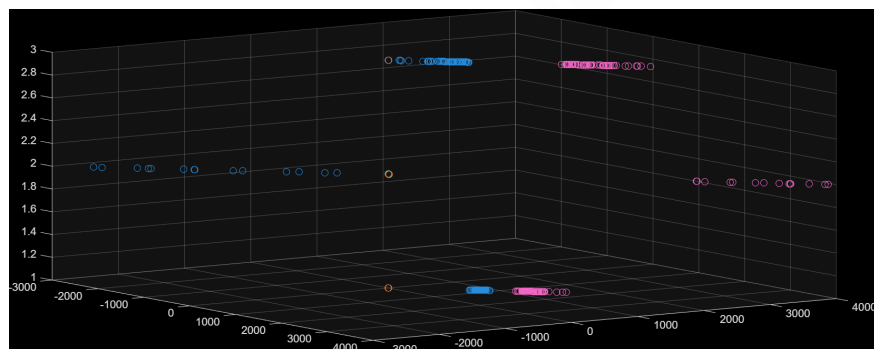


fig. 7 : Figure scatter en 3D de la classification des signaux

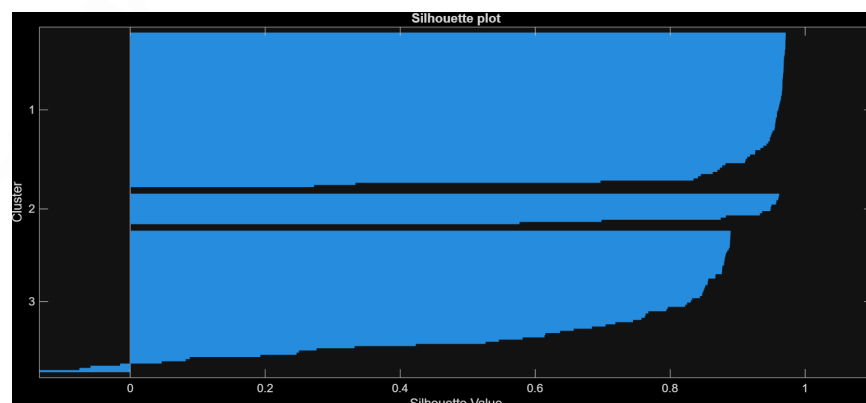


fig. 8 : Figure Silhouette ($k = 4$)

M1 MESI - TP B

Contrairement aux signaux C et F qui sont discontinus et sont présents dans chacun des 3 clusters en petites parties, ce qui confirme bien qu'ils ont une petite partie normale, avec 2 parties majeures qui contiennent des défauts machines.

Sans oublier les signaux B et D, qui sont une autre forme ou type de défaut, ils se situent carrément à l'opposé et sont des signaux contenant un défaut machine plus stable que C et F.

Ensuite, j'ai testé avec $k = 5$, mais on voit que le résultat est très dispersé.

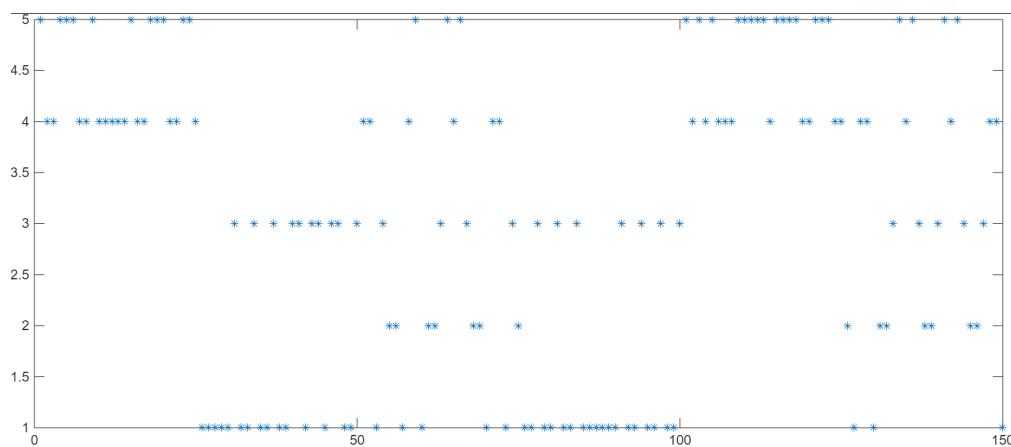


fig. 9 : Figure K-means très instable, très dispersé, peu interprétable

En réponse à la question, je suis convaincu que les signaux A et E correspondent à l'état normal.

Question 4 : Clustering avec distance DTW

Signal A et les 150 fenêtres temporelles



fig. 10 : Figure comparaisons du signal A avec tout le reste

Signal B et les 150 fenêtres temporelles

Daniel BAL

M1 MESI - TP B



fig. 11 : Figure comparaisons du signal B avec tout le reste

Et ainsi de suite pour les autres signaux, on remarque bien que les signaux A et E, sont partout similaires et stables, contrairement aux autres signaux qui sont affectés de défauts machines.

One Class SVM

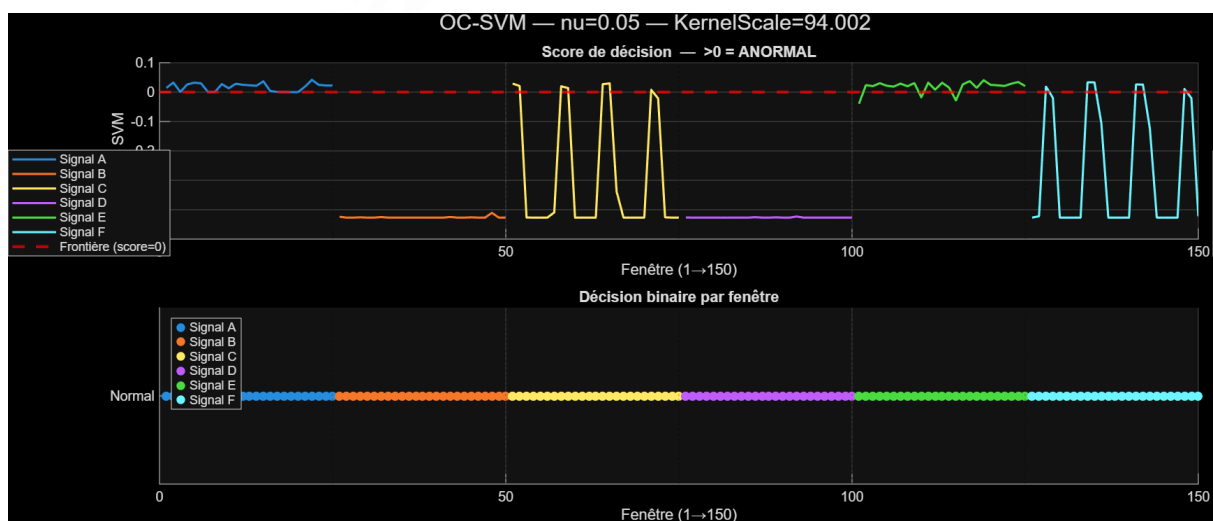


fig. 12 : Figure SVM avec B,D + C, F détecté

M1 MESI - TP B

On observe que les 2 signaux A et E sont sur la frontière (signaux normaux), puis le reste en dehors de cette frontière qui représente les signaux anormaux.

Remarque : On voit des petits pics impulsifs entre l'état normal et anormal des signaux C et F, les roulements abîmés de la machine ou bien des engrenages cassés peuvent être à l'origine ce défaut mécanique par chocs. Et donc les signaux B et D ont un défaut mais plutôt stable sans choc.

Isolation Forest

J'avais déjà entendu M. Knittel que Isolation Forest n'est pas très fiable toujours, et c'est vrai car c'est un peu le seul qui n'a pas détecté les signaux anormaux B et D, tandis qu'il a bien détecté les signaux C et F.

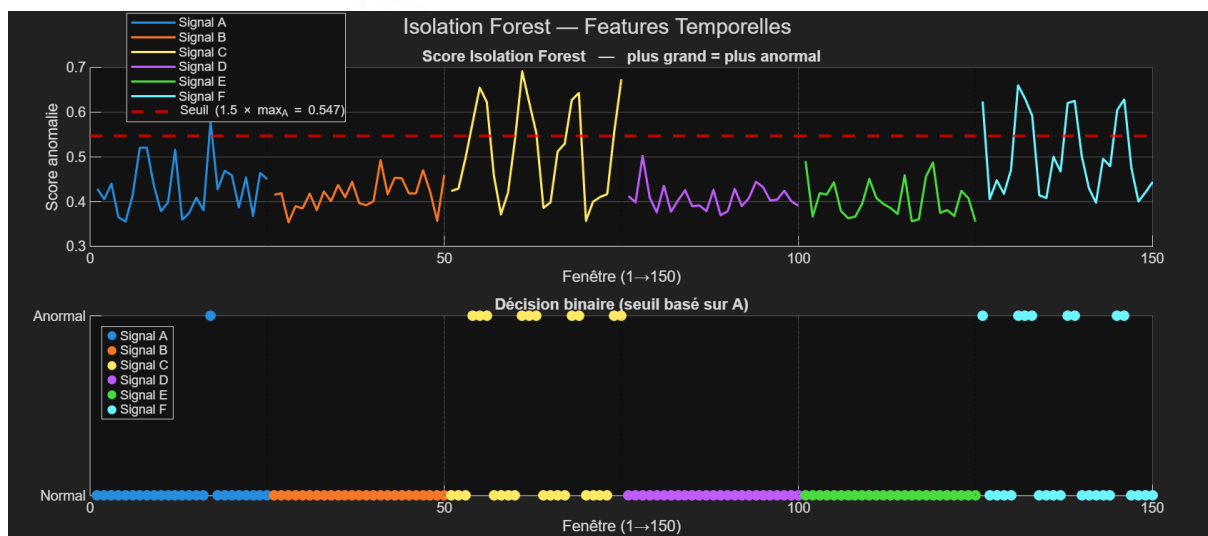


fig. 13 : Figure Isolation Forest avec C et F détecté

On pourrait peut-être essayer en ajustant les paramètres tels que la contamination.

Comparaison de résultats

Signaux	A	B	C	D	E	F
K-means	Normal	Défaut 1	Défaut 2	Défaut 1	Normal	Défaut 2
DTW	Normal	Défaut 1	Défaut 2	Défaut 1	Normal	Défaut 2
SVM	Normal	Défaut 1	Défaut 2	Défaut 1	Normal	Défaut 2
IFOREST	Normal	Normal	Défaut 2	Normal	Normal	Défaut 2
Résultat	Normal	Défaut 1	Défaut 2	Défaut 1	Normal	Défaut 2

Conclusion

Je constate que nous avons pas mal de résultats fiables mais à analyser, en effet cela dépend du modèle utilisé, SVM a l'air plutôt robuste par rapport à Isolation Forest qui est peu fiable. Sinon, K-means aussi est bon pour démarrer mais il ne sait pas trop ce que c'est "normal" donc j'ai dû tout de même prendre du recul.

Le DTW est très bien en visuel, on arrive à voir la forme du signal et réfléchir sur pourquoi un signal est différent des autres, mais il est coûteux en calcul, c'est long.

De plus, le One Class SVM est plus rigoureux, car il apprend explicitement la frontière du signal A qui est normal. Il est juste un peu sensible au réglage de la position de la frontière.

Pour finir, Isolation Forest est plutôt bon pour isoler ce qui est totalement différent, il a pu détecter les défauts sévères, mais il a dû mal à détecter les défauts doux ou les anomalies subtilement différentes des normaux.

Remarque finale : On a déjà fait ce TP en L3, mais là maintenant je comprends mieux l'objectif du professeur, j'ai appris à ne pas faire confiance à une seule

Daniel BAL

M1 MESI - TP B

méthode toute seule, je dois me baser sur plusieurs méthodes ou modèles pour être sûr de ma prédiction.

Ressources

Documentation Matlab

https://fr.mathworks.com/help/?s_tid=user_nav_help

<https://deepfa.ir/en/blog/introduction-to-clustering-algorithms-concepts-applications-key-algorithms>

Rédaction et travail réalisés sans recours à une intelligence artificielle.

Toute imperfection éventuelle relève donc exclusivement de l'intelligence naturelle.